

Università di Macerata

Facoltà di Scienze Politiche - a.a. 2011-2012

Corso di

Statistica

La verifica delle ipotesi sul confronto tra medie

Lezione quindici

Cristina Tortora

cristina.tortora@unina.it



Test per la differenza tra due medie

Esempio (1° caso)

Un automobilista decide di cambiare gli pneumatici della propria automobile, potendo scegliere tra due diverse marche, **X** e **Y**, di uguale prezzo e per i quali le case costruttrici garantiscono una durata media di 35mila km e una deviazione standard di 2000 km, uguale per le due marche.



Prima di effettuare l'acquisto, l'automobilista decide di sottoporre a verifica le ipotesi:

$$H_0: \mu_x - \mu_y = 0$$

$$H_1: \mu_x - \mu_y \neq 0$$

Si informa quindi presso amici e conoscenti che utilizzano i due tipi di pneumatici, trovando 14 persone che utilizzano la marca **X** e 9 che utilizzano la marca **Y**. I risultati sono:

$$\bar{x} = 33.500$$

$$\bar{y} = 36.000$$

Supponendo che la durata degli pneumatici si distribuisca secondo una legge Normale, si verifichi l'ipotesi H_0 con un livello di significatività $\alpha=0,05$.

La stima della differenza fra medie

$$X \sim N(\mu_x; \sigma_x^2) \quad Y \sim N(\mu_y; \sigma_y^2)$$

(varianze note)

$$H_0 : \mu_x = \mu_y \quad H_1 : \mu_x > \mu_y$$

$$n_x; n_y$$

$$H_1 : \mu_x < \mu_y$$

(campioni indipendenti)

$$H_1 : \mu_x \neq \mu_y$$

STIMATORE

$$\begin{aligned} (\bar{X} - \bar{Y}) &\begin{cases} \rightarrow E(\bar{X} - \bar{Y}) = \mu_x - \mu_y \\ \rightarrow \text{Var}(\bar{X} - \bar{Y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y} \end{cases} \end{aligned}$$

Sotto H_0

$$\bar{X} - \bar{Y} \sim N\left(0; \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

Per campioni grandi oppure
quando sia X che Y si
distribuiscono Normalmente

Test per la differenza tra due medie

Esempio (1° caso)

(sulla differenza fra medie)

Un automobilista decide di cambiare gli pneumatici della propria automobile, potendo scegliere tra due diverse marche, **X** e **Y**, di uguale prezzo e per i quali le case costruttrici garantiscono una durata media di 35mila km ed una deviazione standard di 2000 km. Si informa quindi presso amici e conoscenti che utilizzano i due tipi di pneumatici, trovando 14 persone che utilizzano la marca X e 9 che utilizzano la marca Y. Supponendo che la durata degli pneumatici si distribuisca secondo una legge Normale, si verifichi l'ipotesi H_0 con un livello di significatività $\alpha=0,05$.



$$\mu_X = \mu_Y = 35.000 \text{ km} \quad \sigma_X = \sigma_Y = 2.000 \text{ km} \quad X \sim N$$

$$n_X=14 \quad n_Y=9 \quad \bar{X} = 33.500 \quad \bar{Y} = 36.000 \quad z_{\alpha/2} = 1,96$$

• Ipotesi

$$H_0: \mu_X - \mu_Y = 0$$

$$H_1: \mu_X - \mu_Y \neq 0$$

• Significatività

$$\alpha = 0,05$$

• Statistica

$$(\bar{X} - \bar{Y}) \sim N$$

• Regola di decisione

$$\text{Rifiuto } H_0 \text{ se: } \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \right| > z_{\alpha/2}$$

1.

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} = \frac{33500 - 36000}{2000 \cdot \sqrt{\frac{1}{14} + \frac{1}{9}}} = 2,93$$

$$2,93 > 1,96 \Rightarrow \text{Rifiuto } H_0$$

2.

Valori critici

$$\text{non standardizzati: } (\mu_X - \mu_Y) \pm 1,96 \cdot \sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} = \begin{matrix} -1675 \\ +1675 \end{matrix}$$

$$-2500 < -1675 \Rightarrow \text{Rifiuto } H_0$$

Test per la differenza tra due medie

Esempio (2° caso) *(sulla differenza fra medie)*

Un automobilista decide di cambiare gli pneumatici della propria automobile, potendo scegliere tra due diverse marche, **X** e **Y**, di uguale prezzo e per i quali le case costruttrici garantiscono una durata media di 35mila km. Si può ipotizzare che la variabilità della durata sia identica (ma incognita).



Prima di effettuare l'acquisto, l'automobilista decide di sottoporre a verifica le ipotesi:

$$H_0: \mu_x - \mu_y = 0$$

$$H_1: \mu_x - \mu_y \neq 0$$

Si informa quindi presso amici e conoscenti che utilizzano i due tipi di pneumatici, trovando 14 persone che utilizzano la marca **X** e 9 che utilizzano la marca **Y**. I risultati sono:

$$\bar{X} = 33.500$$

$$\bar{Y} = 36.000$$

$$\tilde{s}_x^2 = 4.326.40$$

$$\tilde{s}_y^2 = 3.880.900$$

Supponendo che la durata degli pneumatici si distribuisca secondo una legge Normale, si verifichi l'ipotesi H_0 con un livello di significatività $\alpha=0,05$.

La stima della differenza fra medie

$$X \sim N(\mu_x; ?) \quad Y \sim N(\mu_y; ?)$$

(varianze ignote ma uguali)

$$H_0 : \mu_x = \mu_y \quad H_1 : \mu_x > \mu_y$$

$$n_x; n_y$$

$$H_1 : \mu_x < \mu_y$$

(campioni indipendenti)

$$H_1 : \mu_x \neq \mu_y$$

$$s_x; s_y$$

STIMATORE

Sotto H_0

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\tilde{S}^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}} \sim T_{n_x + n_y - 2}$$

Per campioni grandi oppure
quando sia X che Y si
distribuiscono Normalmente

$$\tilde{S} = \sqrt{\frac{(n_x - 1) \cdot \tilde{s}_x^2 + (n_y - 1) \cdot \tilde{s}_y^2}{n_x + n_y - 2}}$$

La stima della differenza fra medie

Esempio (2° caso) (sulla differenza fra medie)

Un automobilista decide di cambiare gli pneumatici della propria automobile, potendo scegliere tra due diverse marche, **X** e **Y**, di uguale prezzo e per i quali le case costruttrici garantiscono una durata media di 35mila km, potendo ipotizzare che la variabilità della durata sia identica.

L'automobilista si informa quindi presso amici e conoscenti che utilizzano i due tipi di pneumatici, trovando 14 persone che utilizzano la marca X e 9 che utilizzano la marca Y. Supponendo che la durata degli pneumatici si distribuisca secondo una legge Normale, si verifichi l'ipotesi H_0 con un livello di significatività $\alpha=0,05$.



$$\mu_X = \mu_Y = 35.000 \text{ km} \quad \sigma_X = \sigma_Y = ? \quad X \sim N$$

$$n_X = 14 \quad n_Y = 9 \quad \bar{X} = 33.500 \quad \bar{Y} = 36.000 \quad \tilde{s}_X^2 = 4.326.400 \quad \tilde{s}_Y^2 = 3.880.900$$

• Ipotesi

$$H_0: \mu_X - \mu_Y = 0$$

$$H_1: \mu_X - \mu_Y \neq 0$$

• Significatività

$$\alpha = 0,05$$

• Statistica

$$(\bar{X} - \bar{Y}) \sim N$$

• Regola di decisione

$$\text{Rifiuto } H_0 \text{ se: } \left| \frac{\bar{X} - \bar{Y}}{\tilde{S} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}} \right| > t_{\alpha/2; (n+m-2)}$$

Occorre stimare
la varianza comune

$$\tilde{S} = \sqrt{\frac{(n_X - 1) \cdot \tilde{s}_X^2 + (n_Y - 1) \cdot \tilde{s}_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{13 \cdot 4.326.400 + 8 \cdot 3.880.900}{14 + 9 - 2}} = 2038,8$$

$$\tilde{S} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} = \frac{33.500 - 36.000}{2038,8 \cdot \sqrt{\frac{1}{14} + \frac{1}{9}}} = \frac{-2500}{2038,8 \cdot \sqrt{\frac{1}{14} + \frac{1}{9}}} = -2,87$$

$$|-2,87| > 2,08 \Rightarrow \text{Rifiuto } H_0$$

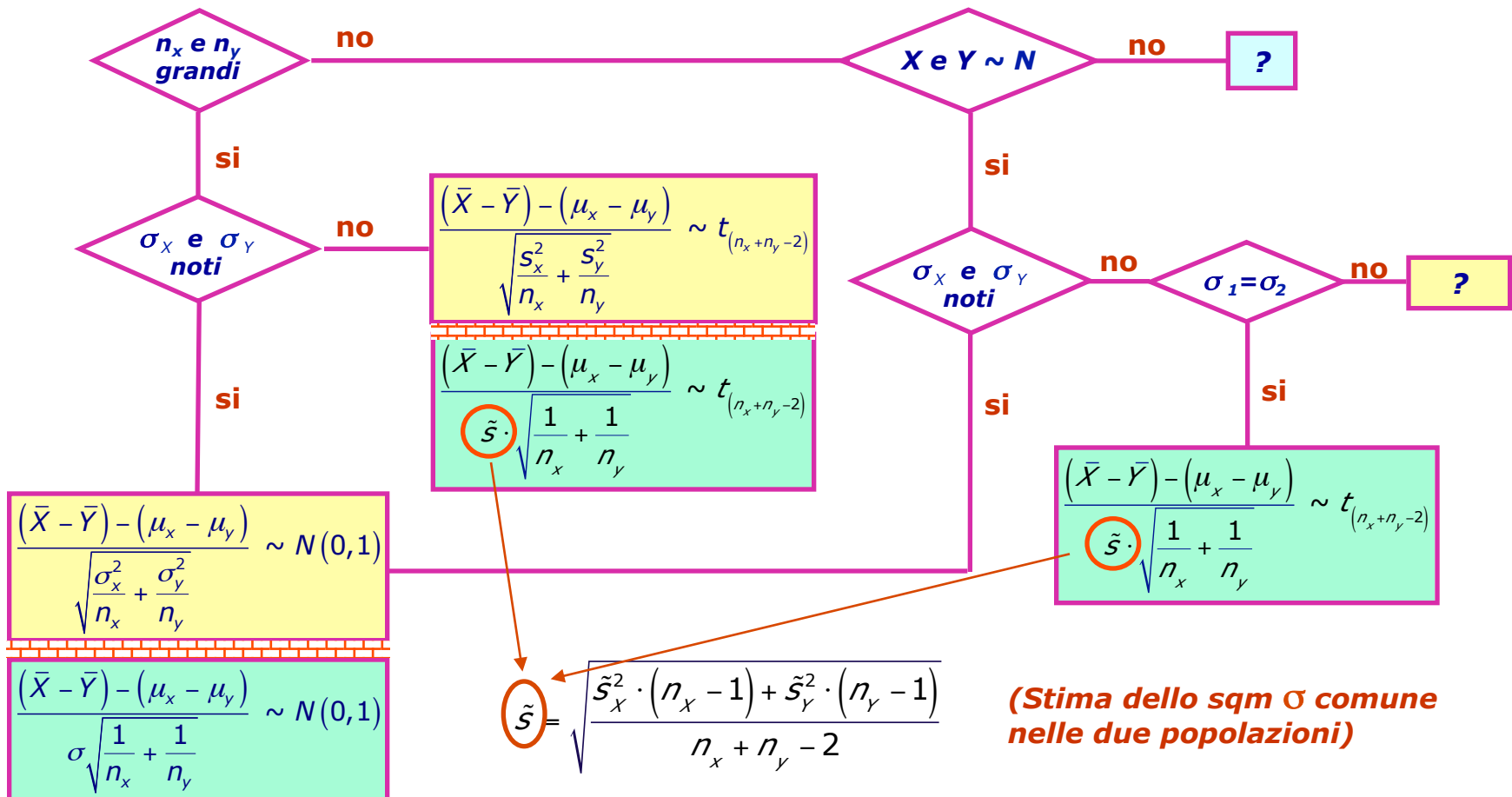
$$t_{0,025;21} = 2,08$$

La stima della differenza fra medie

Riepilogo sulla
stima della differenza fra medie

Eteroschedasticità: $\sigma_X^2 \neq \sigma_Y^2$

Omoschedasticità: $\sigma_X^2 = \sigma_Y^2$



Esercizio

Test sull'efficacia di un trattamento per il controllo della pressione sanguigna con $\alpha = 0,01$. Due popolazioni: la prima formata dagli individui trattati e la seconda formata da individui non trattati:

$$H_0 : \mu_1 = \mu_2 \text{ contro } H_1 : \mu_1 < \mu_2$$

Si suppone che la pressione in entrambe le popolazioni sia una variabile casuale Normale con varianza nota pari a 200:

Due campioni: $n_1 = 15$ individui trattati e $n_2 = 12$ individui non trattati:

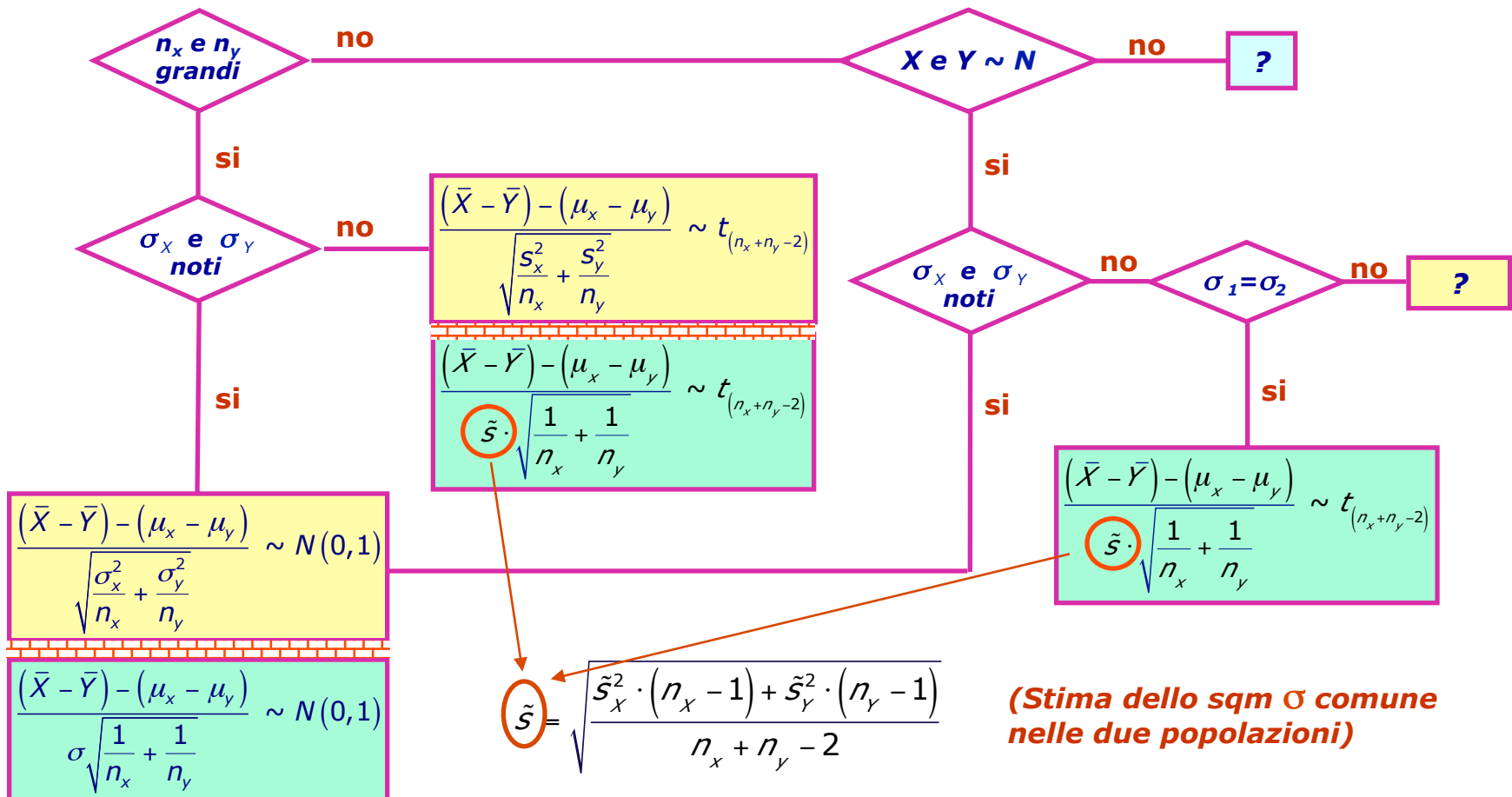
$$\bar{x}_1 = 135,87 \qquad \bar{x}_2 = 170,58$$

La stima della differenza fra medie

Riepilogo sulla
stima della differenza fra medie

Eteroschedasticità: $\sigma_X^2 \neq \sigma_Y^2$

Omoschedasticità: $\sigma_X^2 = \sigma_Y^2$



Soluzione

$$H_0 : \mu_1 = \mu_2 \text{ contro } H_1 : \mu_1 < \mu_2$$

Test per medie – Popolazioni Normali – varianze note

Statistica test e distribuzione sotto l'ipotesi nulla:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

Con σ_1^2 e σ_2^2 che indicano il valore delle varianze nelle due popolazioni e con n_1 e n_2 le dimensioni dei due campioni.

Il valore della statistica test è:

$$z = \frac{135,87 - 170,58}{\sqrt{(200/15) + (200/12)}} = -6,34$$

Ponendo $\alpha = 0,01$ si ottiene il valore della normale standardizzata:

$$-z_{0,01} = -2,326$$

quindi si deve rifiutare l'ipotesi nulla.

Esercizio

Per confrontare l'efficacia di due diverse metodiche, dette A e B, per la misurazione dell'emoglobina si è proceduto nella seguente maniera.

- Si è creato in un laboratorio del sangue artificiale con un contenuto nominale di 15 grammi per 100 cm³ di emoglobina.
- Dal composito sono stati estratti 360 campioni di sangue. Su $n_A=180$ campioni l'emoglobina è stata misurata con la metodica A, mentre per i restanti $n_B=180$ campioni è stata utilizzata la metodica B.
- Sui due differenti campioni si trova:

$$\begin{array}{ll} \bar{x}_a = 15 & \bar{x}_a = 15,02 \\ s_a^2 = 0.046 & s_a^2 = 0.099 \end{array}$$

C'è una differenza significativa tra i contenuti medi dell'emoglobina rilevati con le due metodiche?

Esercizio

$$n_a = n_b = 180$$

$$\bar{x}_a = 15 \quad \bar{x}_b = 15,02$$

$$s_a^2 = 0.046 \quad s_b^2 = 0.099$$

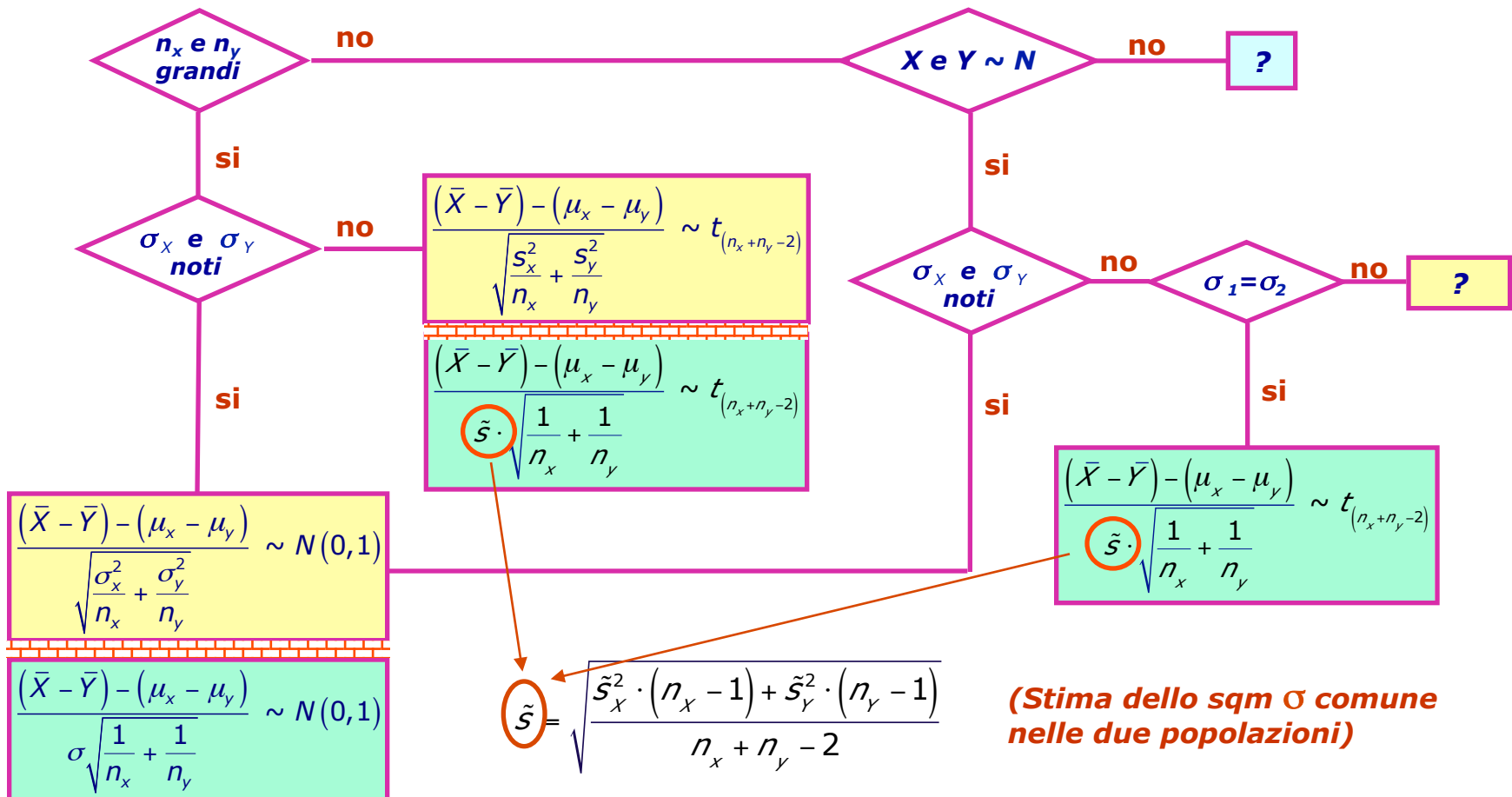
$$H_0 : \mu_1 = \mu_2 \text{ contro } H_1 : \mu_1 \neq \mu_2$$

La stima della differenza fra medie

Riepilogo sulla
stima della differenza fra medie

Eteroschedasticità: $\sigma_X^2 \neq \sigma_Y^2$

Omoschedasticità: $\sigma_X^2 = \sigma_Y^2$



Esercizio

$$n_a = n_b = 180$$

$$\bar{x}_a = 15 \quad \bar{x}_a = 15,02$$

$$s_a^2 = 0.046 \quad s_a^2 = 0.099$$

$$\tilde{s} = \sqrt{\frac{\tilde{s}_x^2 \cdot (n_x - 1) + \tilde{s}_y^2 \cdot (n_y - 1)}{n_x + n_y - 2}}$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \sim t_{(n_x + n_y - 2)}$$

$$\tilde{s} = 0,07719$$

$$t_{n_a + n_b - 2} = 1,96$$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{\tilde{S}_x^2}{n_x} + \frac{\tilde{S}_y^2}{n_y}\right)}} = 7,05$$

Che risultato si otterrebbe utilizzando un approccio p-value?

Esercizio

Supponiamo di voler verificare l'ipotesi che un certo farmaco ABC sia efficace per abbassare la pressione. A tale scopo la pressione sistolica viene misurata su $n=15$ pazienti prima e dopo il trattamento con il farmaco ABC. Si trova media prima è 145 e la media dopo è 123, e $s^2=2.586$. Cosa si conclude?

$$H_0 : \mu_1 = \mu_2 \quad \text{contro} \quad H_1 : \mu_1 < \mu_2$$

$$\bar{x}_a = 145 \qquad \bar{x}_d = 123$$

$$s^2 = 2,586$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\tilde{s} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{(n_x + n_y - 2)}$$

$$T = \frac{\bar{X} - \bar{Y}}{\tilde{s} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = 40,96$$

Esercizio

Si vuole valutare la bontà di un nuovo programma (software) che permette certe analisi di laboratorio sul sangue. E' auspicabile che, a parità di risultati e caratteristiche, il nuovo software sia significativamente più veloce rispetto a quello attualmente in uso.

L'analista sceglie di estrarre due campioni appaiati: quindi lo stesso insieme di campioni di sangue sono testati su entrambi i programmi e viene misurato il tempo richiesto dal vecchio e dal nuovo programma allo scopo di valutare la differenza media fra i tempi necessari per raggiungere i risultati desiderati. Su un campione di $n=10$ provette di sangue, si trova una differenza pari a -0.084 e $s^2=0.08442$. Cosa si conclude?

$$H_0 : \mu_1 = \mu_2 \text{ contro } H_1 : \mu_1 > \mu_2$$

$$\bar{x}_a - \bar{x}_b = -0,84$$

$$n = 10$$

$$s^2 = 0,08442$$

Esercizio

$$H_0 : \mu_1 = \mu_2 \quad \text{contro} \quad H_1 : \mu_1 > \mu_2$$

$$\bar{x}_a - \bar{x}_b = -0,84$$

$$n = 10$$

$$s^2 = 0,08442$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\tilde{s} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{(n_x + n_y - 2)}$$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\tilde{S}^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}} = -6,46$$

Esercizio

In uno studio di marketing, ad un campione di $n_1 = 35$ clienti viene fatto provare un nuovo dentifricio, ed ad un secondo campione di $n_2 = 30$ clienti quello già in commercio. Dopo qualche mese, il grado di soddisfazione medio osservato sul primo campione è $m_1 = 104$ (con $s_1^2 = 56$), mentre sul secondo campione è $m_2 = 102$ (con $s_2^2 = 59$). Si desidera verificare se i clienti preferiscono il nuovo dentifricio rispetto a quello già in commercio.

Esercizio

Per verificare l'efficienza di due processi produttivi (A e B) di una azienda, la produzione (misurata in numero di componenti prodotte) è stata rilevata per un totale di 2000 giorni (1000 per ogni processo produttivo). Si hanno inoltre i seguenti risultati sulle medie e varianze campionarie per i due processi: $m_A=6.15$ e $m_B=9.31$, $s_A^2=1.09$ e $s_B^2=2.15$. Sotto ipotesi di normalità e omogeneità delle varianze, vi sono elementi per sostenere che esiste una qualche differenza tra la produzione media del processo A e la produzione media del processo B?

Esercizio

Si vuole verificare l'efficacia di un trattamento per il controllo della pressione sanguigna. Si esaminano due gruppi: il primo composto da $n_1 = 15$ individui (sottoposto al trattamento), il secondo composto da $n_2 = 12$ individui (trattato con placebo). Sulla base dei seguenti campioni

$x_1 = (155, 28, 130, 140, 106, 158, 130, 150, 145, 135, 128, 146, 122, 130, 135)$

$x_2 = (146, 169, 167, 176, 169, 188, 158, 175, 180, 185, 178, 156)$

si vuole verificare $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 < \mu_2$.

Esercizio

Si vogliono confrontare i tempi medi di risposta degli hard disk di due computer. Si osservano due campioni casuali formati da 130 tempi di risposta e 170 tempi di risposta, rispettivamente, e si trova $m_1 = 10.5$ millisecondi (con $s_1^2 = 8.5$), e $m_2 = 11.7$ (con $s_2^2 = 7.9$). Si vuole verificare $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$.